Human Interaction and Emerging Technologies (IHIET 2023), Vol. 70, 2023, 263–271
https://doi.org/10.54941/ahfe1002951

Operationalizing Ontologies for Competence Management in the Industry

Kurt Englmeier[1] and Pedro Contreras[2]

[1]Schmalkalden University of Applied Science, Faculty of Computer Science, 98574 Schmalkalden, Germany
[2]Cyber Security Centre, WMG, The University of Warwick, Coventry CV4 7AL, UK

ABSTRACT

With the increasing availability of digital resources for on-the-job training, competence management in the industry requires new tools to identify training opportunities for the continuous development of the skills of employees. Our emphasis is to determine which digital courses or further learning resources suit the actual employee's competence in combination with the skills and knowledge she or he aspires to achieve. In this paper, we describe the role of ontologies and, in particular, the ESCO ontology for the development of suitable profiles for learners, learning goals, and learning resources. We describe the matching processes operating on these profiles in order to identify the training opportunities that match best the learner's capacity and aspirations.

Keywords: Recommender systems, Natural language processing, Ontologies, Named entity recognition, Transformers

INTRODUCTION

An important field of competence management is counseling employees who aspire to advance in their careers or want to enhance their skills, competence and knowledge. Together with training opportunities, existing and required skills and knowledge define the setting for the counseling scene. In career counseling, employees expect to learn about their capacities, interests, and goals and how to align them with their personal and career advancement. The employee and probably an interlocutor from competence management work together to identify areas of competence gaps and suitable training programs to reach the aspired learning goals. An employee's capability comprises her or his skills, knowledge, qualifications, and abilities. It can refer to an individual's competence, skills, knowledge, or qualifications, as well as to the resources and technology that are available to a person, organization, or system.

The employee's capabilities and goals and the learning opportunities set the scenery for the design of recommender systems for job counseling, too. First, the system analyses the employees' existing capacities and personal goals. Next, it analyses training opportunities (courses, etc.) and tries to provide personalized information fostering informed decisions about the employee's career development.

When planning the strategy for advancing in our career we expect to get the best advice possible. Searching the web or even consulting the intelligent chatbots, that are in vogue these days, does not help much to avoid the information overload we experience when we use these information sources. Developing a training strategy for advancing in the job or for our personal aspirations often requires a stepwise approach. When we start to plan our strategy, we usually are not in the position to develop a full-fledged goal framework. We need to reflect training possibilities and career options with our capacities in order to fine-tune our learning goal. Only a carefully and thoroughly drafted strategic goal may enable us sound decisions on our future training program. This scenery depicts the leading paradigm for the development of an intelligent recommender system in the context of career development or on-the-job training. We expect such systems to trustworthy guide us along the development of our individually designed training strategy.

Our paper focuses on the design and prototypical implementation of an intelligent recommender system that supports the individual selection of training programs for employees in the industry. Our design approach emerges from the paradigm of cold start recommendation, which is applicable in situations when data on users and items are sparse. The paper outlines generic profile models to summarize capacities and goals of employees as well as training opportunities. These summaries are structured and serve as profiles. We explain the design and operationalization of ontologies for the development of standardized profiles for learners and learning opportunities and the matching processes.

RELATED WORK: COLD START RECOMMENDATION

There are many design qualities trustworthy recommender systems must adhere to, such as fairness-aware, privacy-aware, and transparent recommendations. The pivotal aspect, however, is still accuracy: Do the recommendations accurately match the learners' goals while correctly taking into account their personal capacities?

Recommender systems supporting the development of career strategy planning are basically content-based recommender systems (Pazzani and Billsus, 2007). They try to filter out the learning opportunities that best suit the learner's characteristics and aspirations (see Figure 1). Recommender systems usually make personalized suggestions based on past behavior of user communities (purchases, clicks, ratings, etc.). Commonalities in the behavior of these users are identified and used to predict the intended behavior of a particular user whose goal may fit those of this community. However, the identified community must have a critical size that guarantees accurate reasoning on the community's behavior. Furthermore, data must be available on the past behavior of this particular user that sufficiently overlaps with the behavior of the selected community. If this is not the case, the design of the recommender system better embrace so-called cold start scenarios. There

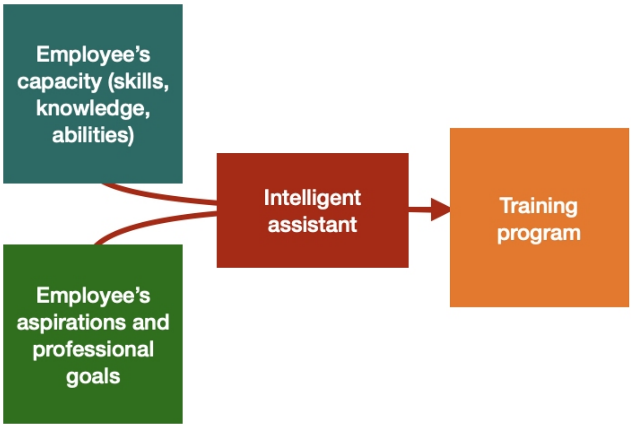

Figure 1: Schema for recommendation in the context of professional and personal formation.

are many suitable approaches for the design of cold start recommendations (Gantner et al., 2010; Mishra et al., 2017; Schein et al., 2002; Huan et al., 2022).

When it comes to recommendations for career strategies, we face a situation that describes user-item cold start, when data are too sparse for a solid user-item recommendation (Zamani and Croft, 2020; Zhang et al., 2014). In other words, we have no historical data available that allows us to identify substantially enough learner communities (users) with similar career strategies (items). Personal careers are too different from each other. Furthermore, privacy-aware systems try to avoid the development of full-fledged user profiles. Privacy concerns, thus, hamper data extraction from personal information such as CVs or job records. A further problem is the lack of substantial data reflecting the career goals of learners. In a cold start situation with such traits, other data that reflect the user are used instead (Barjasteh et al., 2015; Shi et al., 2017). We use job descriptions that may point to the career aspirations the users might have in mind. Job data enable reflections on job opportunities and serve as proxies for possible learner aspirations. With interview-like interactions, the learners are in the position to outline an accurate picture of their aspirations (Nguyen et al., 2007; Zamani and Croft, 2020).

PROFILING TRAINING OPPORTUNITIES AND EMPLOYEE CAPACITIES AND GOALS

Most of the information reflecting the employee's capacities and goals as well as learning opportunities are available in textual form. The recommendation system thus must be in the position to extract meaningful information from texts and to construct suitable (semantic) profiles from the text extractions. This involves natural language processing techniques (NLP) such as text classification and matching. The goal of NLP is to produce profiles for both, the options for the training program (the courses, for instance) and the employee's capacities and goals. A prominent role in this context has named entity recognition (NER), because much information reflecting competences and goals of employees and training opportunities is time-related or otherwise value-related. These data express, for instance, durations, dates, periods of time, or salary. Named entities qualify them in a way that enables their integration with text analysis. NLP helps to distill the essential data from texts, to compose the profiles and to match profiles for users and learning resources.

Profiling starts with the analysis and classification of existing and required capacities of the employee as well as learning opportunities. The learner profiles include goals toward their professional aspirations or the new skill sets the company expects from them. To plan individual training programs for the employees we consider three types of profiles: the actual competence profile of the employee, the goal profile for the aspired training program and the profiles for the training opportunities (learning resources). In fact, we produce one user profile that gradually developed from consolidated job profiles, which, in turn, are derived from job offers related to the same job.

Ontologies help to structure and categorize information extracted from texts and to identify relationships between different pieces of information. An ontology can establish relationships between employees' skills, competences, knowledge, and career goals on the one side and the available training programs on the other. Named entities support the proper equipment of ontologies with relevant information related to time or numerical data such as "2+ years in software engineering". Ontologies start to operate when the annotation of data is completed, which means they start to identify semantically relevant relationships between learner profiles and course descriptions. In the end, the ontologies take care that the system recommends suitable courses to employees. The ontologies incorporate and orchestrate, thus, recommendation models adapted to the individual counseling situation.

We propose, thus, an ontology-based approach for the definition of competence profiles for the individual employee (EP) and formation profiles (FP). EPs can be generated from job descriptions. FPs, that is, profiles of the aspired competence, can be derived from job descriptions, too, and from personal statements of the employees. They reflect the objectives of the training, too. The profiles of the learning resources (LRP) are obtained from course descriptions. In the first step, we develop annotation features using domain-specific BERT embeddings modeled on the European Skills, Competences and Occupations classification (ESCO) system (ESCO, 2023). The models help to produce standardized semantic skeletons for each profile type. Recursive text summarization produces hierarchically structured annotation sets from the descriptions that we link with the semantic skeletons.

OPERATIONALIZATION OF ONTOLOGIES

The operationalization of ontologies in the context of employee training means applying the concepts and relationships defined by the ontology to recommend specific training activities adapted to the individual need of the employee. This can include tasks such as:

- Profiling employee capacities and goals
- Profiling training opportunities
- Matching employee profiles to relevant training courses based on the ontology's representation of relationships between the profiles of employees, courses, and relevant skills or knowledge.
- Facilitating the retrieval of training opportunities by leveraging the ontology's semantic structure.

Profiling means that our system first identifies the most relevant entries of the ESCO ontology (see below) for each kind of description (learner, learner goals, learning resource). The resulting annotations represent the most adequate titles of skills, competences, qualifications, and occupations. The next set of annotations results from the identification of the most relevant terms and named entities for each description. In this context, named entities represent the underlying numerical values by standardized terms, such as "duration" or "amount salary" etc.

Esco Ontologies and Semantic Matching

ESCO describes in detail the skills required for an occupation. It currently provides around 13900 skills linked to 3008 descriptions of occupations. It aims to support job mobility across Europe by improving labor market integration and efficiency. ESCO provides a common vocabulary and ontology (ESCO ontology, 2023) developed around three themes: occupations, skills, and qualifications. In fact, it is based on three types of terms that support language analysis for text matching, namely:

- preferred terms: They are not used for any other occupation or skill and thus are unique. Preferred terms are the ones that best represent an occupation or skill.
- non-preferred terms: They can be synonyms, but also spelling variants, declensions, abbreviations, etc. They are regularly used by jobseekers, employers, or education institutions to refer to concepts related to preferred terms.
- hidden terms: Terms commonly used to address occupations, but are considered outdated, misspelled, or politically incorrect. They are useful for indexing, searching, and text mining purposes, but are invisible to the end users.

As previously mentioned, we aim to fully leverage the ESCO ontology, definitions, and structures to improve the matching of the profiles of learners, job opportunities, and learning resources. This is depicted in Figure 2.

Language and Profile Matching

The profiles are based on different features that form a 'profile template' for each one of the entities addressed. For example, a learning resource (LR) profile can be composed of three relevant features: associated named entities; an associated word vector embedding for LR contents; and another word vector embedding for LR outcomes. As suspected, the more features are available for profiling the richer the profile becomes.

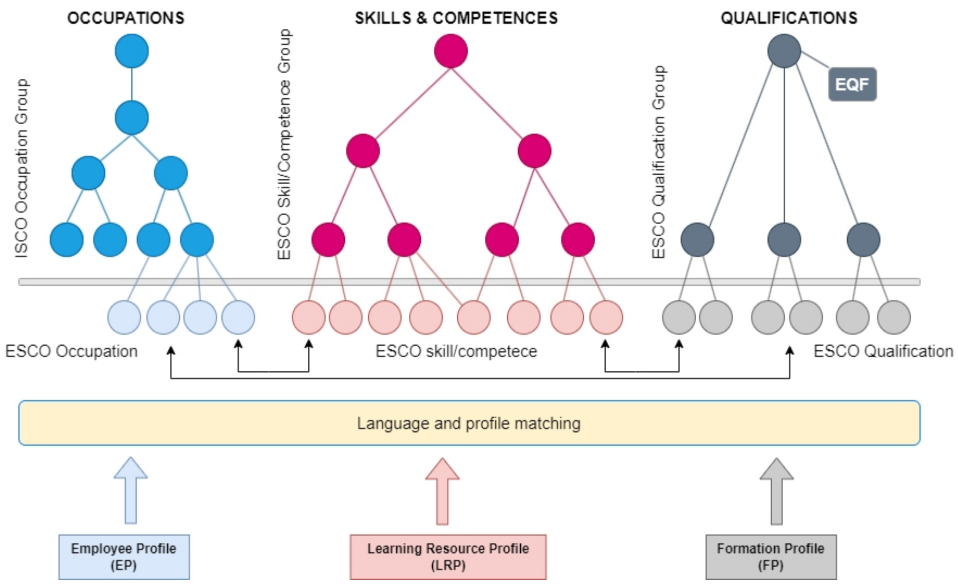

Figure 2: ESCO three pillars and profiles integration, adapted from Vrang et al. (2014) and Shakya (2019).

Here, NLP and word embedding in particular play an important role in semantically matching of profiles. This area of research has had an explosive development in recent years, the old Bag of Word, n-grams, and TF-IDF models for semantic matching have been greatly improved with the arrival of continuous non-contextual (word2vec, GloVe, FastText, etc.) and contextual (Context2Vec, Cove, ELMo, etc.) word representations.

One of the most prominent word embeddings is BERT. Bidirectional Encoder Representations from Transformers (J. Devlin, et al. 2018) is a transformer-based language representation model trained on a large cross-domain corpus. BERT introduces the original concept of masked language model to predict randomly masked words in a sequence, followed by a next-sentence-prediction task for learning the associations between sentences. Through a process of fine-tuning, BERT has achieved state-of-the-art results for a range of NLP tasks.

BERT is very well suited for the task of next-word prediction and when used as a vector embedding in conjunction with a similarity (or dissimilarity) metric, for matching close related semantic similar sentences. In the following section, we describe the use of BERT to match learning resources to the ESCO skills and competences.

MATCHING LEARNING RESOURCES TO ESCO SKILLS & COMPETENCES

We use the ESCO Skills textual descriptions to obtain vector embeddings and the course module contents as queries for matching. We carry out our two initial tests using different BERT embeddings. The first test is as follows:

- We use txtai[1] for embedding and semantic search.
- The sentence-transformers "nli-mpnet-base-v2"[2] model embedding downloaded from HaggingFace was used.
- The sentence-transformers model maps sentences & paragraphs to a 768-dimensional dense vector space and can be used for tasks like clustering or semantic search.
- ESCO Skills descriptions are "embedded" (i.e. mathematically projected) into the sentence-transformers space to create a model and index used for querying
- The embedding process is computationally costly, but it is done only once, to embed new data we can choose to only update the current model.
- Resulting model and index is recorded and used for similarity search based on the k-nearest-neighbor library called faiss (Johnson J. et al., 2021) developed by Facebook.
- Course module contents are used as queries against the embedded model (trained with skills descriptions). We finally retrieve the ESCO skill most closely associated with the query (module contents)

In the second test, we again use the ESCO Skills description to obtain vector embeddings and the course module contents as queries for matching. But in this case, we use a sentence BERT embedding (SBER) (Reimers N., et al. 2019) with a smaller model called "all-MiniLM-L6-v2"[3].

The process is as follows:

- This sentence-transformers model maps sentences and paragraphs to a 384-dimensional dense vector space that can be used for tasks like clustering or semantic search.
- ESCO Skills descriptions are "embedded" using SBERT
- Course module contents are used as queries against the embedded model (trained with skills descriptions), and we retrieve the ESCO skill most closely associated with the query (module contents). The resulting model and index are saved to disk and used for similarity search based on the k-nearest-neighbor library faiss.
- Course module contents are used as queries against the embedded model (trained with skills descriptions), we retrieve the ESCO skill most closely associated with the query (module contents)

The steps above describe a pipeline where embeddings, indexing and matching can be substituted with different methods and algorithms, creating in fact a system that allows comparing different methodologies.

CONCLUSION

In this paper we described the "semantic kernel" of a recommender system supporting competence management. The goal of the system is to find learning resources which suit best the capacities and goals of the learner.

[1] https://neuml.github.io/ txtai/

[2] https://huggingface. co/sentence-transformers/nli-mpnet-base-v2

[3] https://huggingface.co/ sentence-transformers/all-MiniLM-L6-v2

The uniqueness of each learner and each career aspiration forced us to apply the paradigm of cold start recommendation for our system design. The system finally presents a set of job descriptions that address the learner's career aspirations. We take this information as a proxy for the characteristics of the learner's goal. It is the base for the learner's profile that, in the next step, is matched with the profiles representing learning resources.

The paper describes the definition and implementation of the ontologies and shows the first results of their operationalization, that is, the recommendations for individual training programs. It is important to note that NER is applied in this process to improve the expressiveness of the semantic profiles. They complete the definition of the ontologies that represent the individual profiles. NER is useful when it comes to identifying time-related or quantity-related characteristics in job descriptions, such as the number of years a person practices in a particular field, for instance.

The last section of the paper describes two initial semantic matching techniques using transformers to match learning resources against the textual descriptions of skills as defined by the ESCO ontology.

ACKNOWLEDGMENT

Research, as outlined here, is part of the eduplex project under the research framework INVITE of the Federal Institute for Vocational Education and Training (BIBB) funded by the German Federal Ministry of Education and Research.

REFERENCES

Barjasteh, Iman and Forsati, Rana and Masrour, Farzan and Esfahanian, Abdol-Hossein and Radha, Hayder (2015) "Cold-Start Item and User Recommendation with Decoupled Completion and Transduction", Proceedings of the 9th ACM Conference on Recommender Systems.

Devlin, Jacob; Chang, Ming-Wei; Lee, Kenton; Toutanova, Kristina (2018) "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", https://doi.org/10.48550/arxiv.1810.04805

ESCO (2023), European Skills, Competences, Qualifications and Occupations. https://esco.ec.europa.eu/en.

Gantner, Zeno; Drumond, Lucas; Freudenthaler, C.; Rendle, Steffen; Schmidt-Thieme, L. (2010) "Learning Attribute-to-Feature Mappings for Cold-Start Recommendations", Proceedings of the 2010 IEEE International Conference on Data Mining.

Huan, Zhaoxin; Zhang, Gongduo; Zhang, Xiaolu; Zhou, Jun; Wu, Qintong; Gu, Lihong; Gu, Jinjie; He, Yong; Zhu, Yue; Mo, Linjian (2022) "An Industrial Framework for Cold-Start Recommendation in Zero-Shot Scenarios", Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval.

Johnson J.; Douze M.; Jegou H. (2021) "Billion-Scale Similarity Search with GPUs" in IEEE Transactions on Big Data, vol. 7, no. 03, pp. 535-547.

Mishra, Nitin; Mishra, Vimal; Chaturvedi, Saumya (2017) "Tools and techniques for solving cold start recommendation", Proceedings of the 1st International Conference on Internet of Things and Machine Learning.

Nguyen, An-Te; Denos, Nathalie; Berrut, Catherine (2007) “Improving new user recommendations with rule-based induction on cold user data”, Proceedings of the 2007 ACM conference on Recommender systems.

Pazzani, Michael J.; Billsus, Daniel (2007) “Content-Based Recommendation Systems”, in: The Adaptive Web, Brusilovsky, Peter; Kobsa, Alfred; Nejdl, Wolfgang (Eds.), LNCS 4321, pp. 325 – 341.

Reimers, Niels; Gurevych, Iryna (2019). “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks”. Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pages 3982–3992.

Schein, Andrew I.; Popescul, Alexandrin; Ungar, Lyle H.; Pennock, David M. (2002) “Methods and metrics for cold-start recommendations”, Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2002).

Shakya, Aman; Paudel, Subhash (2019). “Job-Candidate Matching using ESCO Ontology”. Journal of the Institute of Engineering, Volume 15, No. 1.

Shi, Lei and Zhao, Wayne Xin and Shen, Yi-Dong (2017) “Local Representative-Based Matrix Factorization for Cold-Start Recommendation”, ACM Transactions on Information Systems (TOIS), Volume 36, No. 2.

The ESCO ontology (2023). The ontology of the taxonomy “European Skills, Competences, qualifications and Occupations”. https://ec.europa.eu/esco/lod/static/model.html

Vrang, Martin le; Papantoniou, Agis; Pauwels, Erika; Fannes, Pieter; Vandensteen, Dominique; De Smedt, Johan (2014), “ESCO: Boosting Job Matching in Europe with Semantic Interoperability,” IEEE Computer, Volume 47, No. 10.

Zhang, Xi; Cheng, Jian; Lu, Hanqing (2014) “Less is More: Sparse Representative based Preference Elicitation for Cold Start Recommendation.” In Proceedings of International Conference on Internet Multimedia Computing and Service.

Zamani, Hamed; Croft, Bruce W. (2020) “Learning a Joint Search and Recommendation Model from User-Item Interactions”, Proceedings of the 13th International Conference on Web Search and Data Mining.